DOCUMENT RESUME

ED 244 736 PS 014 296

AUTHOR Cook, Thomas D.
TITLE Which Questions Do We Need to Ask about Follow Through?: Some Thoughts Stimulated by Considering the Functions of Panel Studies.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE 81
NOTE 22p.
PUB TYPE Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Early Childhood Education; *Program Evaluation; Research Design; *Research Methodology; Research Needs; Research Problems
IDENTIFIERS Evaluation Problems; *Panel Studies; *Project Follow Through; Research Priorities; Research Suggestions

ABSTRACT

National Institute of Education (NIE) research priorities indicate the use of panel measures only during the time children are in Follow Through programs and further seem to indicate that only implementation variables should be measured. This paper raises questions about the desirability of the narrow focus implied by the NIE priorities, the opportunity costs that may be associated with them, the nature of events that seem to have led to the adoption of the narrow focus, and the validity of some of the assumptions on which the focus seems to be based. The paper is divided into four sections. In the first section, key concepts are defined. In the second, concepts are used to describe the place of the NIE evaluation plans within the overall policy space in which evaluations typically take place. Explanations are offered as to why the plans are as they are now. The third section explores how the priorities in the NIE plans lead to a particular sort of panel study. Through consideration of other sorts of panel studies, suggestions are offered concerning how different questions might have been stressed. In the fourth and final section, conclusions are drawn about how the NIE priorities were established; questions are also raised concerning the possibility of expanding research activities within the scope of Strand 1 of the Follow Through evaluation. (RH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED244736

Which Questions do we need to ask about Follow Through?: Some Thoughts stimulated by considering the Functions of Panel Studies[1]

Thomas D. Cook

Northwestern University

PS 014296

[1]This paper was prepared at the request of the National Institute of Education

1981

Introduction

I was asked to write about recent developments in the design and analysis of panel studies, and to relate these developments to the provisional plans that the National Institute of Education (NIE) has proposed for evaluating Follow Through.

The paper uses its manifest purposes to explore a precondition for how panel studies are used. The form of a panel study depends heavily on the specific form of the research questions to be answered. I distinguish three types of panel study, stressing that they can be merged in practice. Repeated measures taken before beginning a program, project, or procedure have most utility for estimating (and extrapolating) ~~motivational~~ maturational trends and group differences in such. Data on this issue helps considerably in making causal inferences. Panel data collected during a program, project, or procedure can describe the nature and context of implementation, can describe levels on outcome variables, and can be used to relate implementation to outcomes. Panel data collected after an experience is over can help estimate the persistence of any changes and help illustrate how any changes are integrated into other aspects of life.

The major thesis of this paper is that NIE research priorities indicate the use of panel measures only during the time children are in Follow Through, and then seem to indicate that only implementation variables should be measured. I raise questions about the desirability of the narrow focus implied by the NIE priorities, about the opportunity costs that may be associated with them, about the nature of the events that seem to have led to this focus being adopted, and about the validity of some of the

assumptions on which the focus seems to be based.

The paper is divided into four sections. In the first we define some key concepts. In the second we use these concepts to describe the place of the NIE evaluation plans within the overall policy space in which evaluations typically take place. We also seek to explain why the plans are how they are. In the third section we show how the priorities in the NIE plans lead to a particular sort of panel study and suggest, by considering other sorts of panel study, how different questions might have been stressed. In the final section, we draw conclusions about how the NIE priorities were arrived at, whether more questions cannot be answered without undue cost increases, whether it is sufficiently realized that one can show that services are implemented well that are of little utility, whether the time frame of Strand I has been utilized fully in exploring whether questions of a different type might be broached, and whether new demonstration projects are needed in Strand I--at least initially.

Some Terminology

What is a Panel Study? Panel studies involve measuring at least two constructs or at least two occasions on the same units. The units are usually individuals, but need not be. Though two waves is a minimum, theorists and practitioners of panel studies are unanimous in stressing the desirability of more waves. I concur with their opinion. However, the resources for multiple waves have to come from resources that could have been used for other evaluation purposes, and so the number of waves influences the total set of research questions that will and will not be asked. The measurement of two constructs is also a minimum, and we will see later that modern "modeling" techniques require more constructs than this.

Panel studies are also characterized by the measurement of persons who have not previously been stratified into different comparison groups. Thus, in studies of television viewing and violence, a panel study would require a large sample of children whose levels of viewing and violence are measured at different times. From this measurement would emerge information about children who differ in viewing. These differences could then be considered as treatment contrasts, even though they were not initially designed for this purpose. (The researcher might, though, have planned the sample selection to ensure wide variability in presumed television viewing, for without such variability many research questions cannot be answered.) The absence of explicitly designated comparison groups (especially no-treatment control groups) makes it possible to do away with a feature of experimental design that costs money, causes political headaches, and sometimes fails to serve its intended baseline function.

The Units about which Evaluation Questions are asked. Some evaluation

questions focus on the program as the unit of study. In the Follow Through case, this would involve phrasing questions about Follow Through (as opposed to not being in Follow Through or being in some program with similar-appearing goals that is aimed at all or part of the same population). Programs rarely die; it is also likely that major decisions about programs depend on political processes, and not on research results. Certainly, it seems from several sources that Congress' intent is not to pass judgments on Follow Through, which is already widely seen as an established service program. For these reasons, inferences at the program level are presumably of little interest to many persons.

Other evaluation questions focus on what I call "project types," a conception that is close to what were called curriculum based "models" in the original Follow Through. With models, the principle aim is to compare different curricula and modes of delivering curricula that have some similar-appearing aims, though they will also differ from each other in their unique aims and in the unique emphases they place on any shared aims. The difficulties involved in comparing models are legion, including (a) how to deal with differences in aims; (b) the political conflicts associated with a hot-house horse race; and (c) the heterogeneity of projects funded from the same model. Also, some theorists of education are convinced that prior evaluations have adequately informed us of the models that are more effective in meeting particular goals. For such persons, "success" depends on getting curricula implemented rather than on getting better curricula designed.

Questions can also be asked about projects. In the Follow Through

context these are the specific sites that have adopted a model sponsor. Typical questions about projects revolve around detecting particularly "successful" projects (hopefully according to a set of heterogeneous criteria); or in relating projects that stress particular activities to changes in particular criteria.

Procedures are also often the focus of evaluations. By procedures I understand activities that occur within projects to achieve subgoals that are presumed to be instrumental in furthering larger goals. In the Follow Through context, procedures of interest might include: How to increase parental involvement; how to increase the time spent on learning tasks, etc. In the NIE materials I was sent these procedures are often referred to as "services."

Evaluations can be designed around other units (e.g., policies, products, personnel, etc.), and an evaluation may involve trying to answer questions about several different units. It should also not be forgotten that evaluations can be designed to discover what are the units most worth asking questions about. Indeed, determining units is one of the more important functions of evaluability assessment or what NIE and ASPE have called "exploratory evaluation in the Follow Through context. It is my understanding that an evaluability assessment of Follow Through has been conducted, and I have consulted a November 1979 summary of the assessment entitled, "Update of the Follow Through Task Force Activities." This update specifies that a decision has been reached to concentrate Follow Through on services, with a 20% effort devoted to research for the purpose of identifying ways of better implementing procedures that are common to

all or many Follow Through projects.

The Types of Evaluation Questions that are asked about any Unit. Irrespective of whether programs, models, projects, procedures (or policies or products) are at issue, most of the specific evaluation questions that are asked, and most of the searching for important research questions that takes place, can be codified as belonging to one of six types of research question.

First come questions about the clients, the audience. In Follow Through, questions of this type refer to the number of children served, their demographic profile relative to what is known of the desired target audience, the numbers of parents who are involved in different ways, the profile of involved parents, etc.

Second come questions about the nature, quantity, and presumed quality of services that are delivered. Issues here concern inferring the educational services that were and were not delivered and also describing the educational context in which the delivery took place, for the context will often facilitate or impede changes in whatever criteria are deemed important. Theory, professional experience, and pilot-testing on site, are usually used to select the particular implementation variables examined. Indeed, all three of these sources are evidenced in NIE's list of possible themes for the Strand I evaluation to emphasize.

A third type of question is about effects. For many commentators on evaluation, this is one of the more crucial and novel aspects of evaluation. In this domain we want to know: How effective is a procedure, project, model, program, etc. in bringing about X or Y for the clients of the program. At issue, here, are both intended and unintended effects, short-term and long-term.

A fourth type of question relates to how a program influences higher-order aggregates that may or may not include members of the client group. I call questions of this type questions about impact, and in the Follow Through case impact questions might refer to effects on families, neighborhoods, other school curricula, other school programs, etc. Impacts are harder to bring about and are more distal than effects on primary target audiences (who, in the Follow Through case, are children).

A fifth type of question that I think useful is about financial costs--total costs, cost per unit per time interval, cost-effectiveness, and--for the adventurous with an underdeveloped sense of the tenuous nature of its assumptions--cost-benefit analysis.

Finally, evaluations often aspire to asking questions about causal process. This is to gain an understanding of processes that mediate particular patterns of implementation, effectiveness, or impact. Such a concern presupposes the utility of differentiating between simple causal relationships (of the form: When I flick the switch the light goes on) and more complex explanatory processes (the light goes on because the current passes along the wire, it strikes a filament in a light, etc.). Knowledge of causal processes helps design more efficient procedures for delivering services than does the more "black box" knowledge associated with identifying dependable causal relationships whose mediating mechanisms are not known well.

Many reasons exist for suggesting the utility of these distinctions I have made about six types of evaluation. One of them will hopefully emerge in the next section.

The Policy Space for the proposed Strand I of the Follow Through Evaluation

Locating the Strand I Evaluation Plans. The NIE documents I have seen presume that the unit for evaluation is the procedure (or service) and not the program, model, project or anything else. They also presume that the major type of question to be asked is about success in implementation. I draw these inferences from the "illustrative themes for pilot projects" contained in the document: "Plans for Follow Through Research and Development" by NIE staff members. The themes require identifying:

- Means to increase instructional time in Follow Through classrooms through improved management of services;
- New patterns of in-service training and selection of teachers to gain better instructional management, including cooperative agreements between schools, teacher training institutions and teacher associations or unions;
- New ways to systematically involve parent and community groups in planning and conduct of Follow Through programs, including the use of parents and families to provide instruction in the home;
- New uses of information systems, including testing and evaluation results, to bring better diagnostic and prescriptive information to bear on Follow Through student learning needs;
- New ways to facilitate support of school building and district administrators for the substantial changes typically required by innovative Follow Through procedures.

The policy space in which it is planned that evaluation of Follow Through should take place is illustrated in the matrix below, where units of evaluation are crossed with the major types of question posed in

evaluation. The one cell entry is where NIE proposes to be. It seems both from the Plans for Follow Through Research and Development and from the summary of the "exploratory evaluation" (or evaluability assessment) that the rest of the matrix will remain unexplored territory.

Table 1. The Place of the proposed Strand I Evaluation Activities in Evaluation Policy Space

Unit of Evaluation --- Question asked in Evaluation	Program	Model	Project	Procedure
How many children of different backgrounds are reached by the program/model/project procedure?				
How well is each of the units implemented?				Strand I of the proposed NIE evaluation
How effective is the unit on direct recipients of its services?				
How impactful is the unit?				
What costs are associated with it?				
Which processes mediate the observed patterns of results?				

Justifying the Strand I Evaluation Plans. The Strand I evaluation plans seem to be the result of an "exploratory evaluation" (evaluability assessment) conducted under the auspices of Joe Wholey when he was in ASPE, together with some employees of USOE. I have had access to a summary of the assessment and of the then Secretary of Education's reaction to it. The discussion below is based on my reading of the summary as well as on background knowledge of the history and politics of Follow Through. My concern is to try to identify the bases on which a decision was made to cast the evaluation as a research project aimed at inferring ways of implementing important procedures or "services."

(a) The Choice to focus on Procedures or "Services." Let us consider why procedures may have been chosen as the unit of evaluation. It may seem particularly fruitless to study Follow Through at the program level when it is already considered to be an established service, has a powerful political constituency to support it, and anyway evaluations rarely influence global decisions about programs.

Alternatively, one could study models. But why do this when there is as much variability (in learning "gains" at least) within models as between them, and when the horse race between models causes political headaches whose public manifestations serve to undermine the credibility of all Follow Through evaluation efforts? Moreover, one might think that one already knows which curriculum-based models are "effective," and major remaining issues are how to get local authorities to sponsor such models and how to get them implemented well when they are adopted.

One could alternatively examine projects. But why evaluate these unless (a) there is a rapid turnover of Follow Through schools at the local

level, or (b) realistic plans exist to expand the number of Follow Through sites and (c) NIE is in a position to affect the curriculum that is implemented in new projects? It would seem that both either (a) or (b) and (c) would have to be true to make worthwhile a focus on identifying successful projects and then examining why they may be successful.

Procedures, on the other hand, are less threatening as units of evaluation. They do not question the core rationale of a sponsor; and they are usually not seen as relevant to major decisions about a project's fate. Also, some procedures are common to a large number of projects or--like parental involvement--are mandated for each project. Research on transferrable successful procedures may have an audience, therefore, particularly since it is easier for projects to change specific procedures than to change educational philosophy. Finally, with the level of background experience we have in evaluating Follow Through, some would argue that we need greater anthropological wisdom about what goes on in Follow Through classes before we ask questions about such grandiose units as programs or models or even schools as projects. One of the greatest sins of the original Follow Through, as we now see with cheap hindsight, was premature grandiosity and an inadequate modesty in the face of reality in general, and of Murphy in particular.

The above argument, which I stress is partly hypothetical, is superficially persuasive. And we shall soon see, however, it overlooks three factors:

- first, one can modify procedures in projects whose basic conception is flawed and from which children benefit relatively little in obvious and important ways;

- second, the time frame assumed for evaluation in Follow Through might well allow a more differentiated evaluation strategy that moves from the evaluation of procedures to the evaluation of other units;
- third, the validity of some of the assumptions in the above argument are murky--at least they are murky to me.

(b) The Choice to focus on Questions of Implementation. We turn now to the decision to focus on questions of implementation, as opposed to focussing on other types of question or as opposed to presuming that the questions worth asking still needed to be discovered. I infer that the implementation of procedures or services--i.e., how well is something done--is at issue, rather than, say, the effectiveness of services--i.e., what effects does a service have if it is conducted well and parents are better involved, students spend more time at learning tasks, etc.--from the following quotation from the OE paper entitled, "Update of the Follow Through Task Force Activities":

> The Follow Through program of the future will have two clear purposes--first, it will provide effective comprehensive services to poor children in elementary schools in the nation; second, it will fund activities designed to improve our understanding of the ways that comprehensive educational services may be most effectively delivered to financially needy elementary school children.

Note that the reference here is to "the ways that services can be effectively delivered" and not to ways in which demonstrably effective services can be effectively delivered.

It is not difficult to see how the decision to focus on implementation may have arisen. After all, effectiveness is usually more difficult to pin down with confidence; and impact is even harder. Moreover, an explicit

interest in effectiveness and impact can lead to political battles, as can cost issues. Besides, it is presumed by some that the current literature indicates that time on task increases performance, and that parental involvement solidifies political bases (among other things). Why, then, should one examine the effectiveness of procedures that can be assumed to be effective? Finally, one could argue that exploring implementation will help, not only in aiding sponsors and adopting schools to implement better, but will also provide powerful clues to help uncover causal mediating processes.

But the same problems exist for a focus on implementation as for a focus on procedures or services. We may well be examining how well procedures are implemented that are not successful, or only successful by certain criteria, or are only successful under a restricted set of conditions. Also, one has to ask whether a focus on the implementation of procedures needs to be exclusive of a focus on the effectiveness of procedures.

Relating Panel Studies to Priorities in Evaluation Questions

Panel studies involve at least two waves of measurement. However, the purposes to which longitudinal measurement is put depends in large part on when the measurement is made. We distinguish three times--before a child enters Follow Through, during Follow Through, and after exiting from it at the end of the third grade. We shall see that the different purposes associated with the different times of measurement speak directly to the issue we have just raised of justifying the dominant research questions.

Measures collected before Follow Through. Two or more waves of measuring achievement, self concept, and background characteristics can serve a very useful purpose if they are collected before a child begins in Follow

Through. Perhaps the major inferential problem that occurs with effectiveness-oriented studies of children is the lack of information about maturational trends and about group differences in such trends. Such trends cannot be sensitively described with only one wave of pre-Follow Through data (i.e., the normal "pretest"). Much more sensitive estimates are possible with two or more pretest waves. Indeed, it is just this feature which provides the rationale for the so-called "dry run experiment."

Collecting data prior to Follow Through is less of a priority the more one focusses on implementation as opposed to effectiveness or impact and on procedures as opposed to projects or models. However, in the Follow Through case, the transfer from Head Start eligibility to Follow Through eligibility may mean that for many children some measures are in their "file." If they are available for enough students, are directly relevant, and are of reasonable quality, then most of the advantages of more than one pre-project measurement wave can be gained. The advantages are not iron-clad guarantees that the observed maturational trends will continue into the future for a specific type of child. Such extrapolation is the crucial untested assumption that may, however, be partially probed using other sources of data. But while two measurement waves prior to Follow Through is no panacea, it is a vast improvement over current practice with a single time of pretest measurement.

Measures collected during Follow Through. Multiple waves of measurement during Follow Through offer the potential for (a) comprehensive description of the services delivered, both in terms of quantity and quality; (b) continuous assessment of performance measures--at the level of school, class, teacher, parent, and child and (c) a chance to relate implementation

and effectiveness measures cross-sectionally and with time lags. A data set with these three characteristics would go some way towards meeting the temporal precedence and covariation criteria of causal inference; and with enough grounded theory and experience and with high quality measurement the data set might plausibly rule out many alternative interpretations to preferred causal inferences.

A description of implementation processes is enhanced by multiple measurement waves, since implementation is usually a dynamic process. To measure at one time would give little sense of the learning and feedback that goes into improving implementation at the local level. Nor would it assess as many of the intermittent outside forces that impinge on implementation to improve or impede it. Moreover, if children are to be measured in terms of the amount and quality of services they receive, this measurement is often better the more it is based on the stable level of services a child receives. Measures that depend on measurement on a single day or week may be unstable because of time-bound factors that happen to increase or decrease the value of observations at the time of measurement. Finally, it should be noted that multiple waves of measurement give the researcher a chance to be a student, to learn which features of the children's, teacher's and parent's experiences deserve to be measured. Later waves can therefore include new constructs that reflect such learning by the evaluators. The current NIE question emphases seem to me to be most relevant to questions that could be answered with multi-wave measurement of implementation <u>during</u> Follow Through.

However, it is also possible <u>during</u> Follow Through to measure the performance of children as well as the activities, etc. in which they

participate. In my reading, the measurement of outcomes plays little role in the current plans for Follow Through. (However, I acknowledge that I have not seen all the relevant documents and that the measurement of performance may be taken for granted because of school testing practices.) Major difficulties with the repeated measurement of performance include political factors--what should be routinely monitored as child-level outcomes, which "interest group" will object to particular measures, and also technical factors--what can we measure well that is likely to change in the time span of Strand 1, and will response formats encourage memory of prior responses?

It is in relating procedures to performance that panel studies seem to many to be most likely to be useful. Alas, though, the state-of-the-art ~~for achieving~~ is imperfect. We know enough not to trust old dogs like cross-lagged panel correlation because nothing can be bought with so few assumptions, as Rogosa and Cook and Campbell have pointed out. We think at present that we need to work ~~with~~ within a framework of structural equation models. To use such "methods" best we also believe that one should postulate many theory-based models of the causal relationship between procedures and outcomes, and should put these into competition with each other as opposed to testing the goodness of fit of only a single model. Moreover, most of us believe that the constructs in these systemically related models should be measured with at least two fallible operations so that inference is at the level of latent constructs (factors).

But some problems remain, and they are tough old ones. One involves how to avoid specification error. Many answer to this that, in the absence of multiply validated and grounded theories, one should take care to put

models into competition with each other, realizing that the exercise is explicitly theoretical in a substantive sense and that, if the theory were so good, one would not need to be doing the modeling! A second problem is how to deal with reciprocal causal influences. Much work is taking place on this ~~theory~~ thorny issue, but I am dubious that break-throughs are near. Some reasonable shots can be made, as our friends in macro-economic theory and methods are continually doing.

Multiple Waves after Follow Through. Follow Through projects are trivial if any initial gains they cause fail to persist after Grade 3. Also, the program, or any project or procedure within it, is less important if any gains are not capitalized upon in a child's later career so that they help him or her in other aspects of school or life outside of school. Apologists for Follow Through might argue that the program is not responsible for maintaining gains and for translating them into better performance in other school or non-school areas. Such factors depend on other factors, most of which are stacked against the kinds of children who are eligible for Follow Through. The apologists are absolutely right in one sense, but may be misleading in another. Programs do not exist in a social void, and if the institutions and programs that a child experiences after Follow Through do not capitalize upon the program, then its overall utility has to be called into question. But such issues aside, the function of follow-up panel measures taken after exiting from Follow Through would be to describe the persistence of changes and to examine how such changes might facilitate other changes in the school or non-school life of a child, parent or teacher.

Measures taken after exiting from Follow Through at the end of the

third grade--or measures taken after "graduating" from some experience within Follow Through--are more obviously related to effectiveness than implementation questions. Moreover, where delayed measures were made in the past, they seem to have been associated more with evaluating programs, models, and projects than with evaluating procedures. This is not an inevitable relationship, though I suspect it is a probabilistic one.

Mixing Times of Measurement. An evaluation can be designed to measure constructs at several times, some of them coming before an experience to be evaluated, others coming during it, and others coming after it. In other words, the design of evaluations permits mixing times of measurement so as to tap into the different strengths of measuring before, during and after.

The current NIE evaluation plan seems to lead researchers to the use of a panel study with multiple waves of measures of how services are implemented, and indicate measurement during the program. I see no a priori need to restrict oneself to implementation measures collected during Follow Through (because performance measures can also be collected then), or to restricting oneself to measures that are only taken during Follow Through.

Conclusions

I have tried to describe the policy space in which it is provisionally planned that Strand I research and development activities should take place. According to plans as I have read them, this space is defined by asking questions about procedures (services) rather than programs, models, projects or anything else, and by a concentration on the quality with which procedures are implemented rather than on how effective, impactful, cost-effective, etc. they are. Such a priority suggests the need for multiple waves of measurement of implementation during Follow Through. I briefly sketched

the advantages of this, but did not dwell on ways of actually conducting the research.

My major concern is not with how to use panel methods to implement NIE's priorities. Rather my concern is with these priorities themselves.

- First, how were they originated? It seems from memoranda I have read that a small OE and ASPE team puzzled through what the qeustions worth asking were in consultation with other OE and Follow Through managers. Later, the then Assistant Secretary of Education narrowed the questions further. Where is the input from the multiple stakeholders that we have come to expect in evaluation? And if such input is now to emerge, how impactful can we expect it to be since tentative plans have already been formulated by powerful Federal groups with relatively homogeneous interests?

- Second I suspect that I could design feasible evaluations to answer a wider range of evaluation questions than I see addressed in the NIE guidelines. Most of the policy space in my Table 1 is, for example, ignored. Moreover, I am immodest enough to believe that I could do this at little, if any, additional cost in terms of either money or the quality of answer to questions about service implementation. Is there to be any chance to expand on the restrictive set of questions in the current evaluation plans? Can the net be cast wider to catch more fish?

- Third, I wonder if there is sufficient cognizance that one can implement well services that are generally ineffective or that are effective only under certain conditions. Do we really believe, for example, that if children spend more time on an ineffectual task, they will learn? Time on task, like parental involvement, the use of media materials and other services mentioned by NIE, presume effective curriculum materials, among other things.

Are we so confident with Follow Through that we know about such effective materials, or that local schools will choose sponsors that have them? To overstate the case, high quality implementation may be necessary for important outcomes but it is not sufficient. Given this, can we study implementation without relating it to effectiveness in the same study or in an immediate follow-up study?

- Fourth, Strand I is long enough that a quasi-programmatic evaluation plan might be developed that begins with a study about both the quality and effectiveness of implementation; and then, in the next study, asks how such procedures can be transferred and whether, once transferred, they affect the crucial overall outcome measures of relevance to Follow Through. (By the latter I do not mean just achievement!)

- Fifth, the NIE documents suggest setting up demonstration projects to examine new patterns of implementation. But what is the rationale for such projects, or for embedding new practices within old projects when considerable variability probably already occurs across projects in each of the procedures of interest (time on task, etc.). Utilizing existing variability might be quicker and cheaper than setting up demonstrations, and would allow one to move on to other questions.

My concern is not to push any particular conception of how Follow Through should be evaluated. It is to raise questions about the questions worth asking about Follow Through. My motive for doing this was that the question of question priorities emerges as soon as one considers panel studies and the different purposes that are usually met by measures collected before, during, or after an experience that is to be evaluated.